Quantifying cycling traffic fluency based on big mobile tracking data

Anna Brauer *[1,2], Ville Mäkinen †[1] and Juha Oksanen ‡[1]

[1]Finnish Geospatial Research Institute FGI, National Land Survey of Finland, 02340 Masala, Finland
[2]Dresden University of Technology, 01069 Dresden, Germany

June 25, 2020

Summary

Activity tracking data collected by mobile applications opens up a new, data-driven perspective on monitoring cycling in the city. In this work, we demonstrate how a large set of trajectories can be used to measure the cyclability of an urban infrastructure. We achieve this by defining the cycling traffic fluency index that describes the smoothness of cycling traffic on segments of a street network. Bias, uncertainty, and the divergence of infrastructure popularity presents challenges to the method, but within these limits, the index could be applied in city planning or as a routing criterion.

KEYWORDS: mobile tracking data, cycling traffic fluency, trajectory analysis, city planning, behavioural pattern inference

1 Introduction

As the awareness of climate change is increasing, cycling as an environment-friendly option is becoming a more and more attractive mode of transportation (Pasha et al., 2016). Cycling has numerous advantages, including low air and noise pollution and positive health effects (Pucher and Buehler, 2012).

To promote biking, a city needs to provide suitable infrastructure, e.g. dedicated cycling lanes (Sener et al., 2009). However, without reliable up-to-date data about the cycling traffic, it is difficult for city planners to decide how the cyclability of a city could be improved most cost-efficiently. Traditionally cycling traffic is monitored by sparsely placed automatic counting devices and manual counting campaigns. Trajectory data from mobile sports tracking applications provides excellent additional source for understanding a population's biking behaviour (Oksanen et al., 2015).

Previous studies show that cyclists dislike stopping and waiting (Stinson and Bhat, 2005; Menghini

*anna.brauer@nls.fi
†ville.p.makinen@nls.fi
‡juha.oksanen@nls.fi

et al., 2010), and prefer to ride at their desired speed without interruptions. We refer to this as traveling fluently. We have studied possibilities to measure cycling traffic fluency (CTF), i.e. the smoothness the cycling traffic flow, based on big mobile tracking data. The method that we have developed extracts different characteristics from the data and aggregates them on segments of the street network. In the final step, the characteristics are combined into a single quantity, the CTF index, that can be used to compare the fluency of cycling traffic in different locations. The resulting data can provide insights into the cyclability of the city. A more extensive presentation of our method and results has been submitted (Brauer et al., *submitted*).

2 Data

The primary data in this work is a set of 50,357 cycling trajectories, recorded with a mobile sports tracking application between 2010–2012, and set public by the application users. All trajectories were recorded in the Helsinki Metropolitan area. The street network data, required for the aggregation, was obtained from OpenStreetMap[1]. For evaluating our results, we use traffic light data provided by the Helsinki Region Infoshare[2].

3 Methods

In the following, we present how our approach extracts stop- and movement-related characteristics from the trajectories. We show how the extracted values are transformed into normalized indices that we finally combine into the cycling traffic fluency index. We also present methods to validate the results.

3.1 Preprocessing

To handle the spatial uncertainty of the trajectories, we apply Gaussian smoothing (Schüssler and Axhausen, 2008) and hidden Markov model-based map matching (Newson and Krumm, 2009), which also maps the trajectories to the street network. We identify and discard parts of the trajectories that do not represent cycling or contain a high amount of noise. From the remaining parts of the trajectories, we extract different characteristics describing the behaviour of the cyclists.

3.2 Stop-related characteristics

As cyclists dislike interruptions, the locations where cyclists are forced to stop are of great significance. We apply a stop detection algorithm that is derived from CB-SMoT (Clustering-Based Stops and Moves of Trajectories, Palma et al., 2008) and map the detected stops to the street network. We define two properties for the street segments: the stop ratio $\widehat{C}_s$, which is the number of stops on the segment divided by the number of trajectories passing through the segment, and the stop duration T_s, which is the average duration of those stops.

[1]https://www.openstreetmap.org/

[2]https://hri.fi/data/en_GB/dataset/helsingin-espoon-ja-vantaan-liikennevaloristeykset

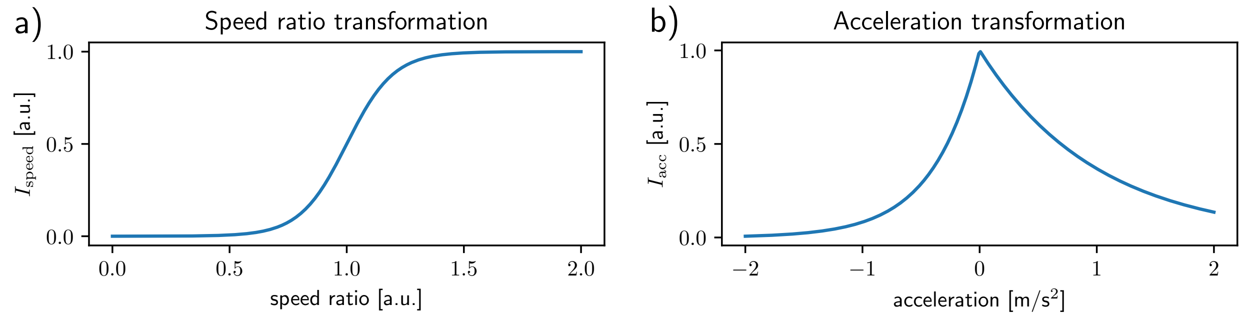


Figure 1: The functions used to transform speed ratio and acceleration into normalized indices I_{speed} and I_{acc}.

The stop duration index $I_{\mathrm{stop\Delta t}}$ is obtained from the segment-wise stop duration values by simple categorization:

$$I_{\mathrm{stop\Delta t}}(s) = \begin{cases} 1 & T_s < 10\,\mathrm{s} \\ 0.8 & 10s \leq T_s < 15\,\mathrm{s} \\ 0.6 & 15s \leq T_s < 20\,\mathrm{s} \\ 0.4 & 20s \leq T_s < 25\,\mathrm{s} \\ 0.2 & 25s \leq T_s < 30\,\mathrm{s} \\ 0.01 & T_s \geq 30\,\mathrm{s}, \end{cases} \tag{1}$$

where T_s is the stop duration of the segment s. The stop ratio is transformed into the stop ratio index using

$$I_{\mathrm{stop\%}}(s) = \begin{cases} 1 & \widehat{C}_s < 0.01 \\ 0.8 & 0.01 \leq \widehat{C}_s < 0.05 \\ 0.6 & 0.05 \leq \widehat{C}_s < 0.1 \\ 0.4 & 0.1 \leq \widehat{C}_s < 0.2 \\ 0.2 & 0.2 \leq \widehat{C}_s < 0.3 \\ 0.01 & \widehat{C}_s \geq 0.3, \end{cases} \tag{2}$$

where $\widehat{C}_s$ is the stop ratio of segment s. The decision for the threshold values of both definitions was aided by examining the data. We combine the stop-related indices into a single index I_{stop}:

$$I_{\mathrm{stop}} = \frac{I_{\mathrm{stop\%}} + I_{\mathrm{stop\Delta t}}}{2}. \tag{3}$$

If either of the constituents is high, i.e. if the average stop time is short or the percentage of cyclists having to stop on the segment is low, the stop index should be at least mediocre. Therefore, the average of the two indices is suitable in this case.

3.3 Movement-related characteristics

Cyclists prefer to ride without stops, but also the travel time is often an important factor. Therefore, we also include the average speed v_s on the street segments in the analysis. Naturally, different cyclists travel at different speeds, and what one cyclist considers as a high speed may be slow for another. To take this into account, we also determine the speed ratio $\hat{v}_s$ of the segments. The speed ratio of a single trajectory traversing a segment is defined as the trajectory's average speed on that segment divided by the mean traveling speed of the whole trajectory. In other words, the speed ratio tries to take the cyclist's preferred speed into account. The speed ratio of a street segment is again the average of the speed ratio of all trajectories passing through that segment.

We also include the acceleration of the trajectories in the analysis. Uninterrupted cycling roughly corresponds to low absolute acceleration values, i.e. to no considerable changes in speed. We note that absolute acceleration values in cycling are low in general and therefore the relative error of acceleration will be higher compared to speed.

The segment speed ratio $\hat{v}_s$ is transformed into the speed ratio index I_{speed} with the equation

$$I_{\text{speed}}(s) = \frac{1}{1 + \mathrm{e}^{-10(\hat{v}_s - 1)}}, \tag{4}$$

where $\hat{v}_s$ is the speed ratio of the segment s (Figure 1a). We use non-linear transformation to enhance the differences between segments with speed ratio close to 1, since the value distribution of the speed ratio is a bell-shaped curve with a small standard deviation.

The acceleration index I_{acc} is defined as

$$I_{\text{acc}}(s) = \begin{cases} \exp(-a_s), & a_s > 0\,\mathrm{m/s^2} \\ \exp(2.5a_s) & \text{otherwise,} \end{cases} \tag{5}$$

where a_s is the average acceleration on segment s. We introduce asymmetry into equation 5 to penalize deceleration more than acceleration (Figure 1b). For both indices, the exact forms of the transformation functions are not critical and can be thought of as parameters of the method, analogously to choosing e.g. the activation function for neural networks (Karlik and Vehbi, 2011).

The movement-related indices are gathered into the movement index I_{move}:

$$I_{\text{move}} = 2\frac{I_{\text{speed}} \cdot I_{\text{acc}}}{I_{\text{speed}} + I_{\text{acc}}}. \tag{6}$$

In this case, we use the harmonic mean so that the speed must be reasonable (high I_{speed}) and the traveling must be smooth (high I_{acc}) for a segment to get a high overall value. The harmonic mean captures this behaviour better than the arithmetic mean.

3.4 Cycling traffic fluency index

Finally, we combine these indices to the CTF index:

$$I_{\text{fluency}} = (1 + \beta)\frac{I_{\text{move}} \cdot I_{\text{stop}}}{\beta \cdot I_{\text{move}} + I_{\text{stop}}}. \tag{7}$$

Figure 2: A part of the street network of Helsinki, overlaid with the stop hot spots. We infer that most stops are caused by traffic lights and street intersections.

We use again the harmonic mean. The weighting parameter β can be used to adjust the relative weights of the two constituents.

3.5 Evaluation

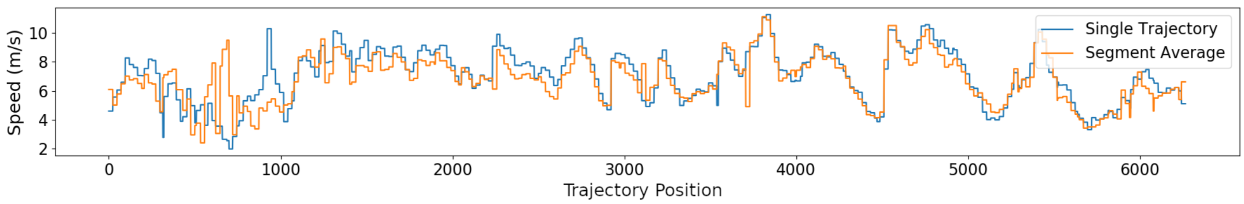
We can assume that in cities, intersections are the most common reason for unintentional stops. Using the DBSCAN algorithm (Ester et al., 1996), we cluster the previously found stops into stop hot spots and analyze their proximity to intersections and to traffic lights. We assign the cause of a stop hot spot to be “traffic light”, if it is closer than 30 m to a traffic light, or else “intersection” if it is closer than 15 m to any intersection. Otherwise, the cause is classified as “other”. In a small field study, we gather reference data regarding the cause of selected hot spots.

To evaluate the explanatory power of the segments’ speed and acceleration values, we compare speed and acceleration of single trajectories to the respective characteristics of the segments traversed by them. For this, we use trajectories that are not part of the original dataset. The correlation is measured using Pearson’s correlation (Rodgers and Nicewander, 1988).

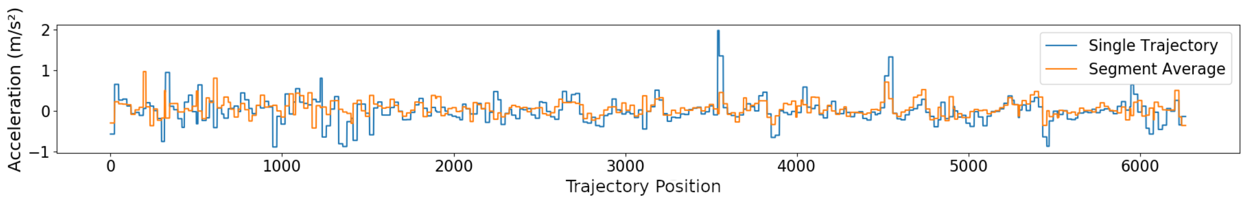
4 Results

According to the stop cause analysis, many stop hot spots are associated with nearby traffic lights or street intersections. 27 % of all hot spots are caused by traffic lights, and 64 % by intersections. A field study confirms the significance of traffic lights as a stop cause, but suggests that the number of hot spots caused by intersections without traffic lights is overestimated. An example from the inner city of Helsinki is shown in Figure 2.

The comparison of segment characteristics with properties of independent trajectories shows excellent agreement regarding the travel speed of most trajectories (Figure 3). There is more variance for the acceleration values, but the correlation is still obvious. This suggests that the segment-wise speed and acceleration values correspond to the on-street circumstances, at least to a certain



(a) Speed, Pearson's $r = 0.810$.



(b) Acceleration, Pearson's $r = 0.333$.

Figure 3: Comparison of speed and acceleration values of an independent trajectory and the corresponding segment averages that are derived from the original dataset.

degree.

Visual analysis of the CTF index shows that in the inner city the average fluency is quite low, but it increases with the distance from the center. Possible explanations for this are the higher density of intersections and higher traffic volume in the center. Intersections usually represent local fluency minimums, which reflects the results of the stop analysis (Figure 4).

5 Discussion & Conclusions

It is worth noting that the data does not cover all segments, therefore a CTF value cannot be assigned to a large portion of the segments. Besides, the dataset is several years old, which may result in local discrepancies compared to the present day. However, they should not have notable effect to the big picture.

Nevertheless, the stop hot spot analysis and the comparison of segment speeds and accelerations to independent cycling trajectories show that the characteristics extracted from the original trajectory dataset are meaningful.

The CTF index map gives a new perspective of the cyclability of the city. It could enable cyclists to plan more comfortable cycling routes, complementing traditional routing criteria like the shortest time and the shortest distance. Data on cycling traffic fluency could also be utilized in city planning, as it could direct the planners' focus to critical locations in need of improvement.

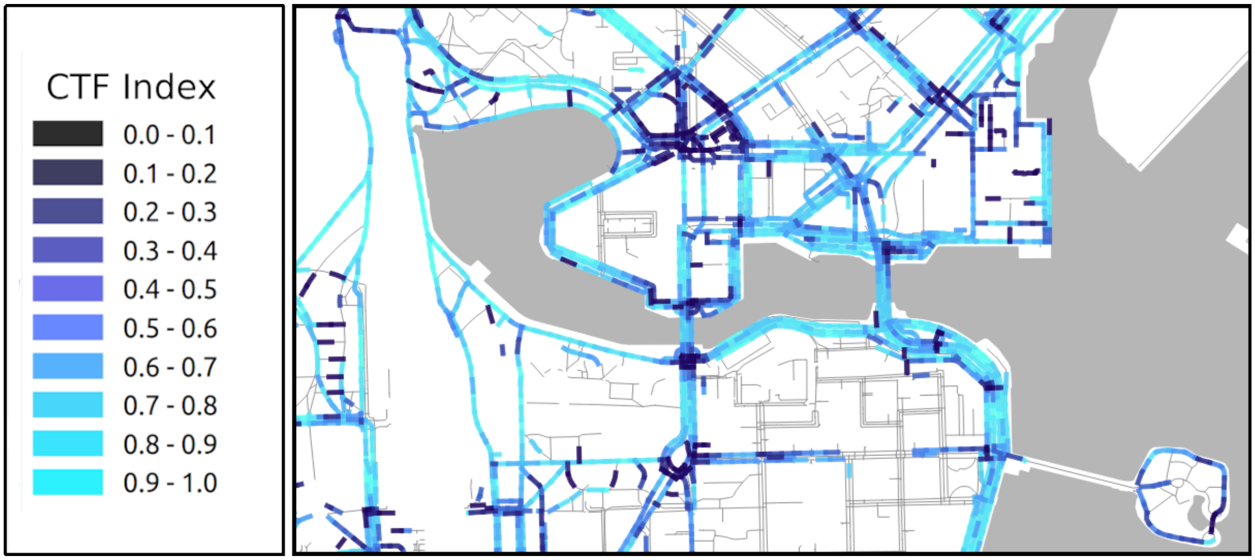


Figure 4: The CTF index values visualized for a part of Helsinki.

6 Acknowledgements

We are grateful to Sports Tracking Technologies Ltd for granting us access to the data used in this work. Only tracks that were originally set public by the users were included in the dataset we received.

7 Biography

Anna Brauer is an assistant researcher in the final year of her computer science studies. At the Finnish Geospatial Research Institute, her work is concerned with mining and analysing human mobility data.

Doctor Ville Mäkinen is a senior researcher in the Finnish Geospatial Research Institute, which is part of the National Land Survey of Finland. His work includes algorithm development and numerical analyses, especially related to digital elevation models.

Professor Juha Oksanen is the Head of the Department of Geoinformatics and Cartography at the Finnish Geospatial Research Institute, which is part of the National Land Survey of Finland. He received his M.Sc. and Ph.D. degrees in geography from the University of Helsinki, where he currently also acts as an Adjunct Professor of geoinformatics. His research interests include cartography, geovisualisation, handling of uncertainty and analysis of large spatio-temporal datasets.

References

Brauer, A., Mäkinen, V., and Oksanen, J. (submitted). Characterizing cycling traffic fluency using big mobile activity tracking data. *Computers, Environment and Urban Systems.*

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231. AAAI Press.

Karlik, B. and Vehbi, A. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122.

Menghini, G., Carrasco, N., Schüssler, N., and Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation research part A: policy and practice*, 44(9):754–765.

Newson, P. and Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 336–343. ACM.

Oksanen, J., Bergman, C., Sainio, J., and Westerholm, J. (2015). Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. *Journal of Transport Geography*, 48:135–144.

Palma, A. T., Bogorny, V., Kuijpers, B., and Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 863–868. ACM.

Pasha, M., Rifaat, S., Tay, R., and de Barros, A. (2016). Urban design and planning influences on the share of trips taken by cycling. *Journal of Urban Design*, 21(4):471–480.

Pucher, J. R. and Buehler, R. (2012). *City cycling*, volume 11. MIT Press Cambridge, MA.

Rodgers, J. L. and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66.

Schüssler, N. and Axhausen, K. W. (2008). Identifying trips and activities and their characteristics from GPS raw data without further information. *Arbeitsberichte Verkehrs-und Raumplanung*, 502.

Sener, I. N., Eluru, N., and Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, U.S. *Transportation*, 36(5):511–539.

Stinson, M. A. and Bhat, C. R. (2005). A comparison of the route preferences of experienced and inexperienced bicycle commuters. In *84th Annual Meeting of the Transportation Research Board, Washington, DC*.